⑲日本国特許庁（JP）　⑪特許出願公開

⑫公開特許公報（A）　昭60－75900

⑤1Int.Cl.⁴	識別記号	庁内整理番号	⑬公開　昭和60年(1985)4月30日
G 10 L 9/08	GLA	7350－5D	

審査請求　未請求　発明の数 1　（全6頁）

⑤4発明の名称　単語音声認識装置

⑳特　願　昭58－183841

㉒出　願　昭58(1983)9月30日

⑦2発明者　下谷　光生　尼崎市塚口本町8丁目1番1号　三菱電機株式会社応用機器研究所内

⑦2発明者　日比野　昌弘　尼崎市塚口本町8丁目1番1号　三菱電機株式会社応用機器研究所内

⑦2発明者　嶋　憲司　尼崎市塚口本町8丁目1番1号　三菱電機株式会社応用機器研究所内

⑦1出願人　三菱電機株式会社　東京都千代田区丸の内2丁目2番3号

⑦4代理人　弁理士　大岩　増雄　外2名

明　細　書

1．発明の名称

単語音声認識装置

2．特許請求の範囲

（1）　入力された音声信号の特徴パラメータを抽出し、該抽出した特徴パラメータと予め登録された複数個の単語音声の特徴パラメータとの類似度を計算して単語音声の認識処理を行なう単語音声認識装置において、

前記認識処理を行なう音声信号区間を規定するために、入力された音声信号の始終端を検出する始終端検出手段を備え、

前記始終端検出手段は、

入力された音声信号の自己相関関数の予め定められた範囲内での最大値CORMAXを計算するCORMAX計算手段と、

前記CORMAXに基づいて、音声信号の始終端の決定を行なう始終端決定手段とを含むことを特徴とする、単語音声認識装置。

（2）　前記始終端決定手段は、前記CORMAXが予め設定されたCORMAXのしきい値を越えたことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第1項記載の単語音声認識装置。

（3）　前記始終端決定手段は、前記CORMAXが予め設定されたCORMAXのしきい値より小さい入力波形が予め設定された区間だけ続いたことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第1項または第2項記載の単語音声認識装置。

（4）　前記始終端検出手段は、入力された音声信号の波形の大きさに対応する量を計算するレベル計算手段をさらに含み、

前記始終端決定手段は、前記CORMAXと前記レベル計算手段の出力とに基づいて、音声信号の始終端の決定を行なう、特許請求の範囲第1項記載の単語音声認識装置。

（5）　前記レベル計算手段は、入力された音声信号のパワーを計算する手段を含む、特許請求の範囲第1項記載の単語音声認識装置。

（６）　前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値を越え、かつ前記ＣＯＲＭＡＸが予め設定されたＣＯＲＭＡＸのしきい値を越えたことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第５項記載の単語音声認識装置。

（７）　前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値を越え、かつ前記ＣＯＲＭＡＸに予め設定された定数を掛けた値がパワーより大きいことを条件として音声信号区間の始端を決定することを特徴とする特許請求の範囲第５項記載の単語音声認識装置。

（８）　前記始終端決定手段は、前記ＣＯＲＭＡＸが予め設定されたＣＯＲＭＡＸのしきい値より小さいことおよび前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値より小さいことの少なくともいずれか一方の条件を満たす入力波形が予め設定された区間だけ続いたことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第５項記載の単語音声認識装置。

（９）　前記始終端決定手段は、前記パワー計算手段で計算されたパワーが予め設定されたパワーのしきい値より小さいことおよび前記ＣＯＲＭＡＸに予め設定された数を掛けた値がパワーより小さいことの少なくともいずれか一方の条件を満たす入力波形が予め設定された区間だけ続いたことを条件として音声信号区間の終端を決定することを特徴とする特許請求の範囲第５項記載の単語音声認識装置。

３．発明の詳細な説明

［発明の技術分野］

この発明は、単語音声認識装置に関し、特にたとえば単語音声区間の始終端検出の改良に関する。

［従来技術］

第１図は従来の単語音声認識装置の一例を示す概略ブロック図である。図において、マイクロホン１１から入力された音声信号は、マイクロホンアンプ１２で増幅された後、ＡＧＣ回路１３に与えられる。このＡＧＣ回路１３は、入力信号の大きさが変動しても、一定出力が得られるように、その内部に備えられた増幅器の利得を自動的に制御する回路である。ＡＧＣ回路１３の出力は、Ａ／Ｄ変換回路１４に与えられ、ディジタル信号に変換される。Ａ／Ｄ変換回路１４の出力は、波形メモリ１５に与えられる。この波形メモリ１５は、１フレームの入力波形データを一時記憶するメモリである。波形メモリ１５の出力は、パワー計算回路２１に与えられるとともに、特徴抽出部３に与えられる。パワー計算回路２１は、波形のパワー（電力）を計算する回路である。パワー計算回路２１の出力は認識処理部６に与えられるとともに、始終端検出回路２２に与えられる。始終端検出回路２２は、音声信号の始終端を検出する回路であり、その出力は認識処理部６に与えられる。一方、特徴抽出部３はディジタルフィルタなどを含んで構成され、入力音声波形の特徴パラメータを抽出する回路である。特徴抽出部３の出力は認識処理部６に与えられる。この認識処理部６には、入力パターンメモリ４および登録パターンメモリ５が接続される。入力パターンメモリ４は、単語音声の認識モードにおいて、特徴抽出部３で分析抽出された認識すべき音声の特徴パラメータを一時記憶するメモリである。登録パターンメモリ５は、登録モードにおいて、分析抽出された登録語の特徴パラメータあるいは標準音声の特徴パラメータを予め記憶するメモリである。認識処理部６は、たとえばマイクロプロセッサやマイクロコンピュータなどを含んで構成され、入力パターンメモリ４と登録パターンメモリ５内の特徴パラメータを用いて認識処理を行なう回路である。このような単語音声認識装置においては、音声区間をフレームと呼ばれる一定時間間隔に分割してフレームごとに音声の特徴抽出が行なわれる。

次に、第１図の回路の動作を説明する。マイクロホン１１から入力された音声信号はマイクロホンアンプ１２，ＡＧＣ回路１３，Ａ／Ｄ変換回路１４を通って一旦波形メモリ１５に記憶される。

特徴抽出部３は波形メモリ１５から１フレーム分の波形データを受取り特徴パラメータの抽出を行なう。得られた特徴パラメータは、登録モードにおいては登録パターンメモリ５に記憶される。一方、認識モードにおいては、得られた特徴パラメータは一旦入力パターンメモリ４に記憶され、その後認識処理部６でパターンマッチング等の手法により認識処理が行なわれる。

一方、始終端検出回路２２は、パワー計算回路２１が計算する音声信号のパワーにもとづいて、音声信号区間の始終端を検出する。認識処理部6-は、この始終端検出回路２２で規定される区間の音声信号を認識すべき音声信号として認識処理を行なう。

第２図は音声信号のパワー波形を示す図である。この第２図を参照して、第１図に示す始終端検出回路２２の動作を説明する。始終端検出回路２２は、音声信号のパワーが予め設定されたしきい値Ｐｓを越えると音声信号の始端を検出し、パワーが予め設定されたしきい値Ｐｅ以下であるフレームが予め設定された区間Ｋｔｈだけ続くと音声信号の終端を検出する。この例では、ｋ１とｋ２がそれぞれ音声信号の始端フレームと終端フレームである。前述のように、認識処理部６は始終端検出回路２２で規定される音声信号区間すなわちｋ１～ｋ２の区間を認識すべき単語音声として認識処理する。したがって、このような認識装置においては、音声信号の始終端検出の性能が認識結果に大きな影響を与える。

第３図は音声に騒音が加わった場合の音声信号のパワー波形を示す図である。この波形の音声信号区間は第２図に示すように、ｋ１～ｋ２であるにもかかわらず、従来装置の始終端検出方法ではｋ３～ｋ４が音声信号区間であると検出する。このように、従来の音声認識装置は、騒音が強い環境下においては、単語音声の始終端検出が正確に行なわれず、認識性能が下がるという欠点があった。

［発明の概要］

この発明は、上述のような従来の装置の欠点を除去するためになされたもので、単語音声の始終端を予め設定された範囲内の自己相関関数の最大値（以下ＣＯＲＭＡＸと称す）を用いて検出することにより騒音が大きい環境下においても始終端検出を正確に行ない得て、認識性能の優れた音声認識装置を提供することを目的とする。

［発明の実施例］

第４図はこの発明の一実施例を示す概略ブロック図である。図において、この第４図の実施例は、以下の点を除いて第１図の回路と同様であり、相当する部分には同様の参照番号を付しその説明を省略する。この第４図の実施例では、ＣＯＲＭＡＸ計算回路２３が設けられる。このＣＯＲＭＡＸ計算回路２３は、たとえば乗算器や加算器からなる自己相関器を含んで構成され、波形メモリ１５から入力される音声信号波形のＣＯＲＭＡＸを計算する。計算されたＣＯＲＭＡＸは、始終端検出回路２２０の一方入力に与えられるとともに、認識処理部６に与えられる。始終端検出回路２２０の他方入力には、パワー計算回路２１の出力が与えられる。すなわち、この実施例の特徴は、パワー計算回路２１で計算されたパワーとＣＯＲＭＡＸ計算回路２３で計算されたＣＯＲＭＡＸとに基づいて、音声信号の始終端を検出することである。

次に、ＣＯＲＭＡＸについて説明する。１フレーム分の波形データをｘ（ｉ）、（ｉ＝１，２，…Ｉｆ）とするとパワーＰは次式（１）で表わされる。

$$P=\sum_{i=1}^{If} x(i)\cdot x(i) \qquad \cdots(1)$$

τ次の自己相関係数ＣＯＲ（τ）は次式（２）で表わされる。

$$COR(\tau)=\sum_{i=1}^{If-\tau} x(i)\cdot x(i+\tau) \qquad \cdots(2)$$

ＣＯＲＭＡＸを求めるために設定した自己相関関数の区間を次数τｓ～τｅ（τｓ～τｅ）とするとＣＯＲＭＡＸは次式（３）で表わされる。

$$CORMAX=MAX[COR(\tau)] \qquad \cdots(3)$$

パワーの大きさが同じ波形であっても母音などのピッチ性の強い波形はＣＯＲＭＡＸは大きく、白色雑音に近い環境騒音などの波形はＣＯＲＭＡＸ

は小さい。この発明は、このことを利用して音声信号の始終端検出を行なうものである。

第５図は音声に騒音を加えた場合のＣＯＲＭＡＸ波形を示す図であるが、図示のようにＣＯＲＭＡＸ波形では騒音の影響が緩和されている。したがって、第１図の回路と同様にしきい値弁別を行なって始終端を検出した場合、音声信号区間はｋ５～ｋ６となり、従来の装置に比べて正確に始終端の検出を行なうことができる。この発明では、ＣＯＲＭＡＸのみに基づいて単語音声の始終端を検出するようにしてもよい。しかしながら、始終端の検出の要素として、ＣＯＲＭＡＸだけでなく音声信号のパワーやレベルなど波形の大きさに対応する量を組合わせて用いると、さらに正確な始終端の検出が行ない得る。そこで、第４図の実施例では、始終端の検出の要素として音声信号のパワーと、ＣＯＲＭＡＸとを用いている。

すなわち、第４図の実施例では、音声信号のパワーが予め設定されたしきい値Ｐｓ以上でしかもＣＯＲＭＡＸに予め設定された定数Ｃｓを掛けた値がパワー以上であるフレームを音声信号の始端フレームとし、パワーが予め設定されたしきい値Ｐｅ以下であるかＣＯＲＭＡＸに予め設定した定数Ｃｅを掛けた値がパワー以下であるかの少なくとも一方を満足するフレームを無音フレームとし、この無音フレームが予め設定されたフーム数Ｋｔｈだけ続くと、音声信号の終端を検出し、最初の無音フレームを音声信号の終端フレームとしている。

次に、第４図の実施例のさらに詳細な動作を説明する。パワー計算回路２１およびＣＯＲＭＡＸ計算回路２３は、波形メモリ１５から１フレーム分の波形データを受取り、それぞれ、第（１）式および第（２）式の計算を行ない、パワーとＣＯＲＭＡＸを始終端検出回路２２０に送る。始終端検出回路２２０は、パワーとＣＯＲＭＡＸを用いて音声信号の始終端判定を上述の方法によって行ない、その結果を認識処理部６に与える。認識処理部６では、始終端検出回路２２０によって検出された始端から終端までの間の音声信号を入力パターンメモリ４あるいは登録パターンメモリ５に格納し、認識処理を行なう。なお、その他の動作は、第２図に示す従来装置と同様である。

なお、他の実施例として、始終端検出回路２２０における始終端の検出は、パワーが予め設定されたしきい値Ｐｓ以上でかつＣＯＲＭＡＸが予め設定されたしきい値ＣＭｓ以上であるフレームを始端フレームとして検出し、パワーが予め設定されたしきい値Ｐｅ以下であるかＣＯＲＭＡＸが予め設定されたしきい値ＣＭｅ以下であるか少なくとも一方を満足するフレームを無音フレームとし、この無音フレームが予め設定されたフレーム数Ｋｔｈだけ続くと終端を検出し最初の無音フレームを音声信号の終端フレームとするようにしてもよい。この場合も第４図の実施例と同様の効果を奏することはもちろんである。

さらに他の実施例として、始端を検出する場合、パワーの値が１フレーム前のパワーの値よりも大きいという条件を加えてもよく、この場合は始端検出能力を向上することができる。

また、上述の実施例では、始終端を検出するための１要素として音声信号のパワーを用いるようにしているが、このパワーに代えてその他音声信号の波形の大きさを表わす量（波形のレベルなど）を計算して始終端検出のための要素として用いるようにしてもよい。

また、上述の実施例では、パワーが予め設定されたしきい値Ｐｓ以上でかつＣＯＲＭＡＸに予め設定された定数Ｃｓを掛けた値がパワー以上であるフレームを音声信号の始端フレームとしたが、始端フレームはこの近傍のフレームにしても差し支えない。

また、上述の実施例では、音声信号の終端フレームを最初の無音フレームとしたが、終端フレームはこの近傍のフレームでも差し支えない。

さらに、上述の実施例では、説明の都合上単語音声認識装置を特定話者登録型としたが、予め標準音声の特徴を登録パターンメモリに登録している不特定話者用の単語音声認識装置であってももちろんよい。

［発明の効果］

以上のように、この発明によれば、ＣＯＲＭＡＸに基づいて単語音声の始終端を検出するようにしたので、騒音が大きい環境下でも音声信号の始終端の検出を正確に行なうことができ、音声認識装置の認識性能を高めることができる。

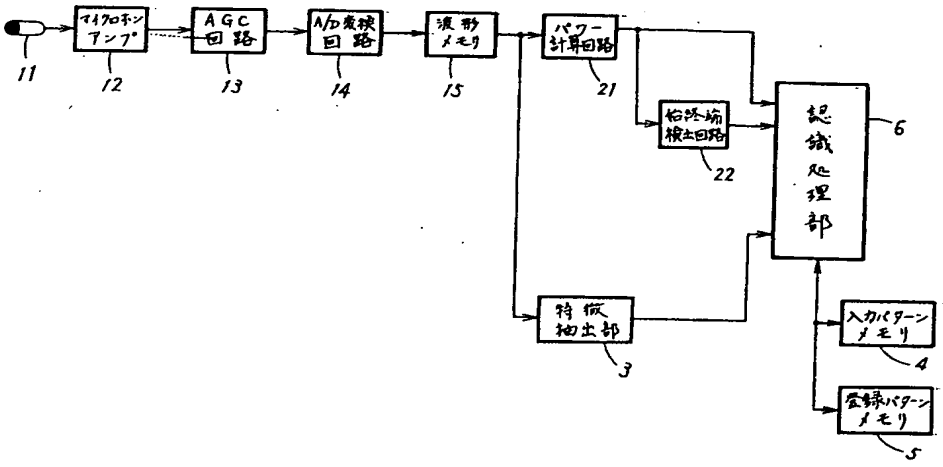
４．図面の簡単な説明

第１図は従来の単語音声認識装置の一例を示す概略ブロック図である。第２図は音声信号のパワー波形を示す図である。第３図は音声に騒音が加わった場合のパワー波形を示す図である。第４図はこの発明の一実施例を示す概略ブロック図である。第５図は音声に騒音を加えた場合のＣＯＲＭＡＸ波形を示す図である。

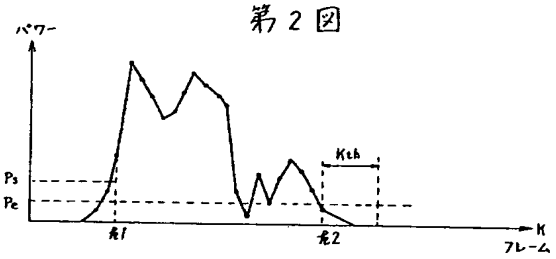
図において、３は特徴抽出部、４は入力パターンメモリ、５は登録パターンメモリ、６は認識処理部、１１はマイクロホン、２１はパワー計算回路、２３はＣＯＲＭＡＸ計算回路、２２０は始終端検出回路を示す。

代　理　人　　大　岩　増　雄

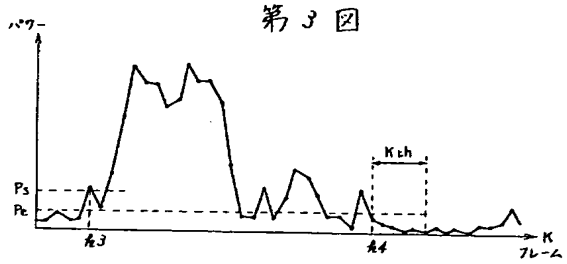
第１図



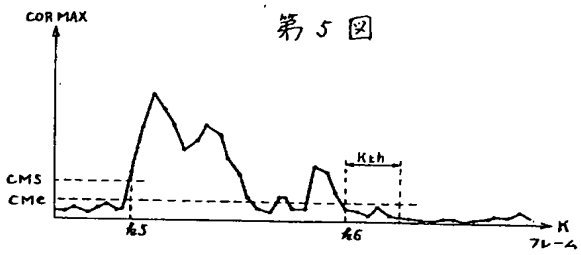
第2図



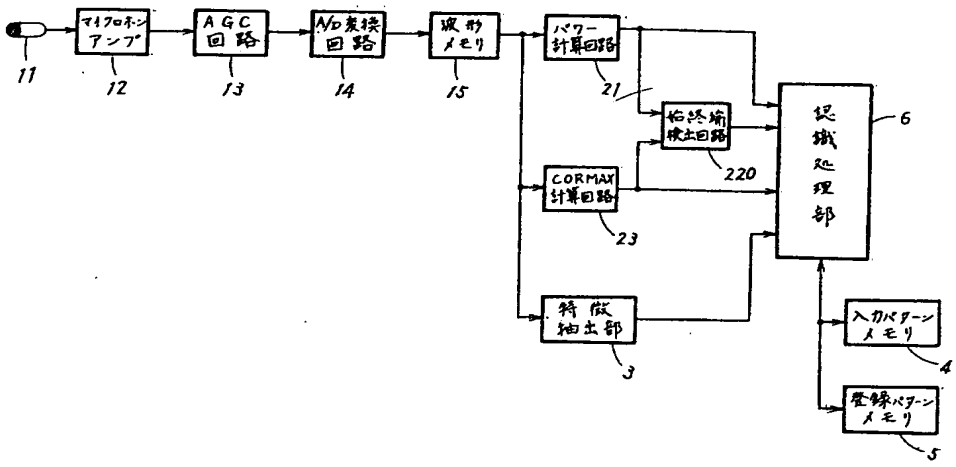
第3図



第5図



第4図



PARTIAL TRANSLATION OF JAPANESE UNEXAMINED PATENT PUBLICATION (KOKAI) NO. S60-075900

Title of the Invention: A Speech Recognition Apparatus
Publication Date: April 30, 1985
Patent Application No.: S58-183841
Filing Date: September 30, 1983
Applicant: MITSUBISHI DENKI K.K.

2. Claims

(1) A speech recognition apparatus which extracts feature parameters of an input audio signal, and calculates a similarity between the extracted feature parameters and feature parameters of speech segments registered in advance, thereby to recognize a speech, the apparatus comprising

a beginning and terminating end detection means for detecting beginning and terminating ends of the input audio signal in order to define an audio signal region targeted for recognition,

wherein the beginning and terminating end detection means includes:

a CORMAX calculation means for calculating a CORMAX, which is a maximum value of an autocorrelation function of the input audio signal in a predetermined range; and

a beginning and terminating end determination means for determining beginning and terminating ends of the audio signal based on the CORMAX.

(2) The speech recognition apparatus according to Claim 1, wherein the beginning and terminating end

determination means determines a beginning end of the audio signal region on condition that the CORMAX exceeds a preset threshold thereof.

(3) The speech recognition apparatus according to Claim 1 or 2, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform with the CORMAX below a preset threshold thereof has lasted for a preset duration.

(4) The speech recognition apparatus according to Claim 1, wherein:

the beginning and terminating end detection means further includes a level calculating means for calculating a quantity corresponding to an amplitude of the input audio signal waveform, and

the beginning and terminating end determination means determines the beginning and terminating ends of the audio signal based on the CORMAX and an output of the level calculating means.

(5) The speech recognition apparatus according to Claim 1, wherein the level calculating means includes a means for calculating a power of the input audio signal.

(6) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a beginning end of the audio signal region on condition that a power resulting from calculation by the power calculating means exceeds a preset threshold thereof, and the CORMAX exceeds a preset

threshold thereof.

(7) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a beginning end of the audio signal region on condition that a power resulting from calculation by the power calculating means exceeds a preset threshold thereof, and a value resulting from multiplication of the CORMAX and a preset constant is larger than the power.

(8) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform has lasted for a preset duration when the input waveform meets at least one of a requirement that the CORMAX is below a preset threshold thereof and a requirement that a power resulting from calculation by the power calculating means is below a preset threshold thereof.

(9) The speech recognition apparatus according to Claim 5, wherein the beginning and terminating end determination means determines a terminating end of the audio signal region on condition that an input waveform has lasted for a preset duration when the input waveform meets at least one of a requirement that a power resulting from calculation by the power calculating means is below a preset threshold thereof and a requirement that a value resulting from multiplication of the CORMAX and a preset number is smaller than the power.